Shchukina, A., Kaźmierczak, M., Kasprzak, P., Davy, M., Akien, G., Butts, C., & Kazimierczuk, K. (2019). Accelerated acquisition in pure-shift spectra based on prior knowledge from 1H NMR. *Chemical Communications*, (64), 9563-9566. https://doi.org/10.1039/c9cc05222d

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Accelerated acquisition in pure-shift spectra based on prior knowledge from ^{1}H NMR†

Alexandra Shchukina,[a] Magdalena Kaźmierczak,[b] Paweł Kasprzak, [a,c] Matthew Davy, [d] Geoffrey Akien, [e] Craig P. Butts, [d] and Krzysztof Kazimierczuk*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Pure-shift NMR enhances spectral resolution, but the optimal resolution can only be obtained at the cost of acquisition time. We propose to accelerate pure-shift acquisition using optimised 'burst' non-uniform sampling schemes [I. E. Ndukwe, A. Shchukina, K. Kazimierczuk, C. P. Butts, *ChemComm*, 2016, 52, 12769] and then reconstruct the undersampled signal mathematically. Here, we focus on the reliability of this reconstruction depending on the sampling scheme and present a workflow for the sampling optimization. It is ready to be implemented in routine measurements and yields a great improvement in reconstruction of challenging cases.

Pure-shift NMR[1,2] techniques allow the suppression of homonuclear *J*-couplings in an NMR spectrum. This gives a boost in resolution as multiplets in a spectrum collapse into singlets. The techniques rely on echo-type pulse sequences with pulses that select a subset of nuclei in the sample to refocus their *J* evolution. The echo is followed by an acquisition of a chunk of a free induction decay (FID) signal. The *J*-couplings, which evolve far slower than chemical shifts, are thus suppressed if the chunk is short enough that they are not resolved over that time period. Sequential acquisition of these short chunks then enables the whole FID to be constructed from these chunks. Pure-shift techniques fall into two types, real-time and interferogram, based on how the chunks are acquired. In real-time experiments, acquired FID chunks alternate with gaps during which the receiver is off and couplings are being refocused. The gaps are excised from the data in order to reconstruct the final FID, leading to an artificially shortened FID and thus real-time experiments lead to spectra with broadened singlets. On the other hand, interferogram acquisitions resemble a standard two-dimensional measurement, where each chunk is measured in a separate increment. The one-dimensional FID is then reconstructed in a pseudo-2D manner from these increments (see e.g. Fig.2 in[3]). Such experiments, e.g., PSYCHE pulse sequence[4], tend to yield spectra of better resolution, but are much more time-consuming. This leads to our goal to accelerate interferogram pure-shift experiments with non-uniform sampling (NUS), obtaining the maximum possible resolution while avoiding reconstruction artifacts in the resulting spectrum.

NUS is applied in various types of NMR experiments to decrease the acquisition time by omitting some measurement points in the FID and reconstructing them afterwards[5,6]. Gaps in the FID result in artifacts which resemble false spectral peaks. A reconstruction algorithm is thus required to separate the true peaks from these artifacts, suppress the artifacts and reveal the underlying clean spectrum[7]. There are several groups of reconstruction methods: linear prediction[8], maximum entropy[9], compressed sensing[10], etc. In follow-up to our previous papers[3,11,12], we chose the compressed sensing (CS) approach. CS achieves the reconstruction assuming that the spectrum is sparse, i.e., the peaks occupy a small part of a spectrum. This is often the case, especially in multidimensional experiments. Out of all the possible values of the omitted points, a CS algorithm selects the set which suits the sparsity assumption best. A number of CS algorithms exist, including CLEAN[13], iterative soft thresholding (IST)[5,14] and iterative re-weighted least squares[5]. All of them solve the task of "cleaning" the spectrum iteratively. Other methods of NUS reconstruction have been discussed elsewhere[15]. In this report we apply the IST algorithm for the reconstructions and also present a comparison to linear prediction results.

The NUS schedule applied in the direct dimension of a pure-

[a] *Centre of New Technologies, University of Warsaw, Banacha 2C, 02-097 Warsaw, Poland. E-mail: k.kazimierczuk@cent.uw.edu.pl*
[b] *Faculty of Chemistry, Warsaw University of Technology, Noakowskiego 3, 00-664 Warsaw, Poland.*
[c] *Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland.*
[d] *School of Chemistry, University of Bristol, Cantock's Close, Clifton, BS8 1TS Bristol, United Kingdom.*
[e] *Department of Chemistry, Lancaster University, Lancaster, LA1 4YB, UK*
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

shift experiment, however, is crucially different from the one used to sample indirect dimensions. Here, one omits whole chunks, not separate points of the FID (Fig.2). We call this type of sampling "burst sampling"[3] and it is more challenging for NUS reconstruction algorithms, as any sampling regularity increases undersampling artifacts[10]. Despite this, however, CS reconstruction has been successfully applied to the burst sampling case[3], and it has recently been shown that it can provide certain benefits for multidimensional NMR as well[16].

In this paper we propose a way to optimize a burst sampling scheme for interferogram pure-shift acquisition based on the information from a conventional (non-pure-shift) ^{1}H spectrum. Our aim is to minimize the risk of high undersampling artifacts and thus assist the reconstruction. There are numerous examples of NUS sampling optimization in literature[17,18], however, to our knowledge, none of these are applied to the challenging case of burst-sampled NUS.

The acquisition in the experiments under consideration is normally carried out in chunks of equal length (see e.g. Fig.2 in[3]), and the chunk length is limited by the anticipated maximum value of J-couplings (shorter chunks are needed for stronger couplings). At this stage we wish to limit possible sampling schemes and so restrict ourselves to simply omitting arbitrary chunks and not changing their length. The number of chunks to be omitted is then selected based on the desired experiment time that one can afford.

Let us look deeper into how exactly a sampling scheme influences the artifacts in the spectrum. The FT of the sampling scheme itself, also known as a point-spread function (PSF)[15], is crucial here. An example is presented in Fig.1 of ESI †. In a spectrum obtained by the FT of the zero-augmented undersampled FID, any spectral peak will be convoluted with the PSF. The convolution, in its turn, will result in every peak having similar patterns of artifacts arising from it. The artifacts from the different peaks co-add, amplifying in some spectral points and cancelling out in others. As the positions of peaks vary from spectrum to spectrum, a single chosen scheme may lead to strong artifact overlap for one spectrum, but not generate any overlap in another (see Fig.2 in ESI †). Thus, to prevent an overlap of artifacts and real peaks, a given spectrum will have its own optimal sampling scheme.

In order to identify the optimal sampling scheme for a given pure-shift spectrum, we propose to use the prior knowledge provided by the conventional ^{1}H spectrum. This allows us to approximately anticipate the frequencies of the peaks that will exist in the as-yet unmeasured pure-shift spectrum through simple multiplet analysis and integration. Crucially, we can also identify the regions where no peaks will appear. With this information in hand, one can choose a sampling schedule most advantageous for the reconstruction, i.e., a scheme optimized to have most artifacts appearing in empty regions of that particular spectrum. It should be noted that we only apply this preliminary knowledge for the sampling optimization, but not for the reconstruction process, as e.g. in SIFT method[19].

We propose the following work-flow for the sampling schedule optimization for a 1D (pseudo-2D) interferogram pure-shift experiment:

1. Measure a conventional ^{1}H spectrum.
2. Perform multiplet analysis to identify approximate frequencies of real peaks.
3. Simulate the expected pure-shift spectrum with singlets on the positions of the multiplets (including those which the analysis failed to classify).
4. Apply different sampling schemes to the simulated pure-shift FID. Take the FT to get the convolution of the simulated spectrum with the PSF of the sampling scheme (spectrum with artifacts).
5. Among all the simulated spectra with artifacts, select the best one (see below for a discussion of possible criteria for designating "best").
6. Measure the pure-shift experiment with this sampling schedule.
7. Perform the CS reconstruction.

In the case of a small number of chunks, it is possible to compare all the possible sampling schedules within a reasonable time (tens of seconds on a standard PC). Otherwise, random sets of schedules can be chosen and the "best" selected from these. For practicality, we chose 10^4 as a limiting number of schedules which might be compared: if there are less than 10^4 possible schedules, they are all probed, if not, 10^4 random ones are chosen. In all cases the schedule where all chunks are gathered in the beginning of the FID, corresponding to simple truncation instead of NUS, should be included as one of the random sample. The first chunk should also always be present in every sampling scheme.

For each sampling schedule, the convolution of its PSF with the simulated spectrum of singlets is calculated and normalized against the simulated singlet spectrum so that the height of the highest peak in both is the same. Then, we propose two options. The first one is to determine the highest absolute deviation between the fully sampled and convoluted undersampled simulated spectra. Then we choose the sampling which yields the minimum highest deviation. Let us call this option the "maximum artifact" criterion. The second option is calculating the ℓ_1 norm of the residual, i.e., the sum of absolute values of the differences between the simulated spectra: fully sampled and undersampled (convolved with PSF). Let us call this method a "sum criterion". We can also think of it as calculating the norm of the artifacts in the 1st iteration of CS reconstruction. The best sampling will be identified as the one with the minimum sum.

Both criteria are inspired by how the CS reconstruction works. The maximum artifact criterion matches the concept of the maximum incoherence of a measurement matrix, while the sum criterion minimizes the ℓ_1 norm of the reconstruction procedure input[10].

The workflow described in this work has been wrapped up in a single package allowing to run it in a fully automated way under Bruker TopSpin software (3.5 and higher). The package is available in ESI. For further updates see `nmr.cent.uw.edu.pl` → `Downloads`.

To see how these two options work, we measured full pure-shift spectra and undersampled them artificially according to the "best" and "worst" samplings of both criteria, reconstructed

and compared to the original. We measured the ^{1}H and pure-shift interferogram spectra of several samples: α-asarone, D-glucose, (R)-(+)-limonene, L-menthol, α-pinene, quinine, sucrose, β-estradiol, all 50mM solutions, and a mixture of 0.5M (R)-(+)-limonene and 50mM α-pinene. We acquired their spectra with the PSYCHE pulse sequence on Varian 700MHz spectrometer equipped with HCN room-temperature probe. The data can also be downloaded from `nmr.cent.uw.edu.pl/downloads`, section "Pure Shift" $->$ "Pseudo-2D pure shift sampling optimizer". We carried out the undersampling and reconstruction with Python scripts available in the ESI †. They exploit *nmrglue* library[20] and *mddnmr* software[21] (the latter can be downloaded from `mddnmr.spektrino.com`). For the spectrum reconstruction, we use the iterative soft thresholding (IST) reconstruction algorithm[5,14,22]. It is based on the ℓ_1-norm minimization of the sought-for spectrum and, at each iteration, picks frequencies which have intensities in the spectrum above a defined threshold and adds them to the output (see a detailed description in[7]). The results for the β-estradiol sample (maximum criterion) are presented in Fig.1. Fig. 2 visualises the best sampling scheme for this case: it applies the best sampling to the simulated FID (step 3 of the workflow). The results for the same sample, sum criterion, are given in the ESI† in Figs.43-45.

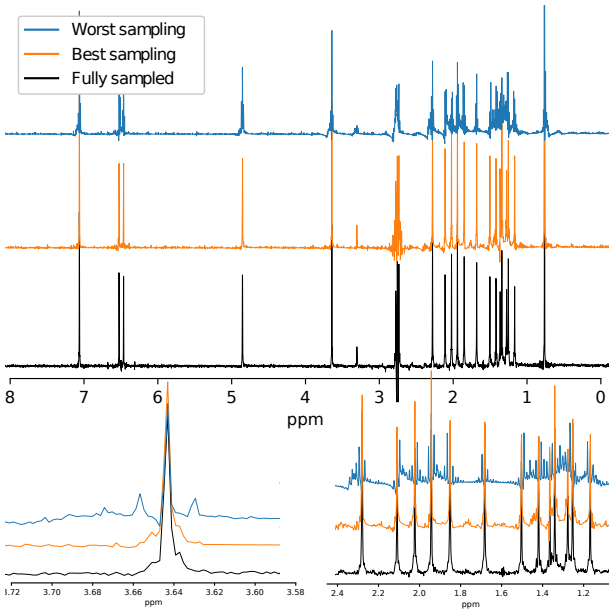

Fig. 1 β-estradiol, 50mM in CD_3OD. Maximum criterion: reconstructions for the worst sampling, the best sampling (both would correspond to acquisition time of 2.5 minutes) and the original fully sampled pure-shift spectrum (13 min 10 s). See the results for other samples, as well as sum criterion results, in ESI†.

In all the cases, we undersampled the FID to take 18 chunks out of 93 (19% sampling). This corresponds to 2.5 minutes of measurement time instead of 13 minutes 10 seconds for the given parameters. In the case of a two-dimensional spectrum (see below), however, the same undersampling level will result in an impressive reduction in time to 1 hour 16 minutes instead of 7 hours 27 minutes. To see how lower sampling levels work, we carried out the same procedure for two 1D examples: β-estradiol and glucose, for 17, 16,... etc. chunks. The results are presented in the ESI† in Figs.47-54.

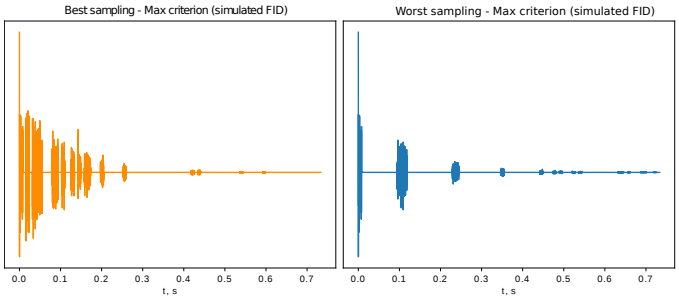

Fig. 2 Maximum criterion: best and worst samplings applied to simulated FID of singlets (point 3 of the workflow).

For the studied examples, we find that the two criteria both give good performance. Across all the samples, the maximum artifact criterion is more likely to leave small artifacts still remaining in the spectrum, while peak heights are almost always perfectly represented. The sum criterion, on the other hand, leads to small deviations of peak intensities or (sometimes, as in Fig. 27 in ESI†) loss of resolution, but is better in terms of suppressing artifacts between peaks. So in general the "maximum criterion" seems more appropriate (and we make it default in our software), but for mixtures with a high dynamic range of peak intensities, where small artifacts may be confused with real peaks, one might choose the sum criterion. We would also note that the sum criterion often gives the best outcome by simply truncating the FID, i.e., only using the first few chunks of the FID. This is not always the case, as is illustrated in Fig. 56 of the ESI†, but is often. So, it is likely that for spectra with very sharp natural linewidths the sum criterion will be even less effective - although IST reconstruction is more robust than linear prediction for extrapolation of truncated FIDs (see Fig.46 in ESI†). It was shown by Hyberts et al[23], where the authors state that IST can extrapolate signals up to 4× and, contrary to LP, does not require not require prior assumptions regarding the number of components.

On the other hand, the difference in reconstruction performance between various sampling schemes is very pronounced. The effect of the sampling scheme is demonstrated here for the β-estradiol spectrum presented in Fig.1 (other examples can be found in the ESI † Figures 3-42). The best sampling scheme based on the maximum artifact criterion gives a reconstruction (orange) which is almost indistinguishable from the fully sampled PSYCHE spectrum (green). As can be seen in Fig. 2, the FID for the best sampling scheme is heavily weighted towards including early chunks but also incorporates a small number of later chunks ensuring that spectrum resolution is appropriately captured and can reasonably be reconstructed without completely relying on IST to extrapolate all of the later data points. On the other hand, reconstruction of the 'worst' sampling scheme (blue in Fig. 1) gives a spectrum which has little resemblance to the fully sampled PSYCHE in some regions.Unsurprisingly, the undersampled FID of the worst sampling scheme (blue in Fig. 2) shows that the chunks are more weighted towards the later, more noisy, portions of the FID.

The optimization procedure is robust in terms of multiplet analysis quality, which is illustrated in Figs.58-63 of the ESI †.

We considered only sampling schemes with fixed and equal chunk and gap lengths. The best samplings worked out very well even within these restrictions. However, there is no practical limitation for varying chunk length and/or gap length if needed. Thus, if one deals with very challenging samples it is likely that the sampling scheme can be improved even further.

The proposed method can be especially useful for interferogram pure-shift 2D experiments with homodecoupling in the F2 dimension (i.e., pseudo-3D experiments). Such experiments are highly time-consuming, reaching up to tens of hours. Thus, omitting some F2 FID chunks can be particularly beneficial. To illustrate this, we performed a 2DJ-ZQS-PSYCHE experiment[24]: a variant of a J-resolved experiment with a pure-shift F2 dimension. The F2 dimension is acquired here in a pseudo-2D manner, thus leading to the pseudo-3D acquisition all together. We carried out a full measurement with the β-estradiol sample (see Fig.3a), as well as two measurements with "best" and "worst" NUS in F2. The best and worst samplings were yielded by the maximum criterion for the one-dimensional case. Afterwards, we performed the IST reconstruction of the NUS datasets. The results are given in Figs. 3b and 3c for the best and the worst sampling schemes, accordingly. One can see an apparent advantage of the "best sampling" case, whose reconstruction is almost identical to the fully sampled data, over the "worst sampling" case, where the reconstruction failed to suppress numerous artifacts in the F2 dimension. The gain on the experimental time is significant: 1h 16min for NUS acquisition instead of 7h 27min for full acquisition.

In summary, the use of the coupled ^{1}H spectrum as a source of prior knowledge to enable the constraint of burst sampling scheme selection for NUS interferogram-style pure-shift NMR experiments provides definite improvement in the performance of CS reconstruction of the corresponding spectrum. We propose the maximum-artifact criterion to find the sampling scheme giving the optimal spectrum quality for the particular sample.

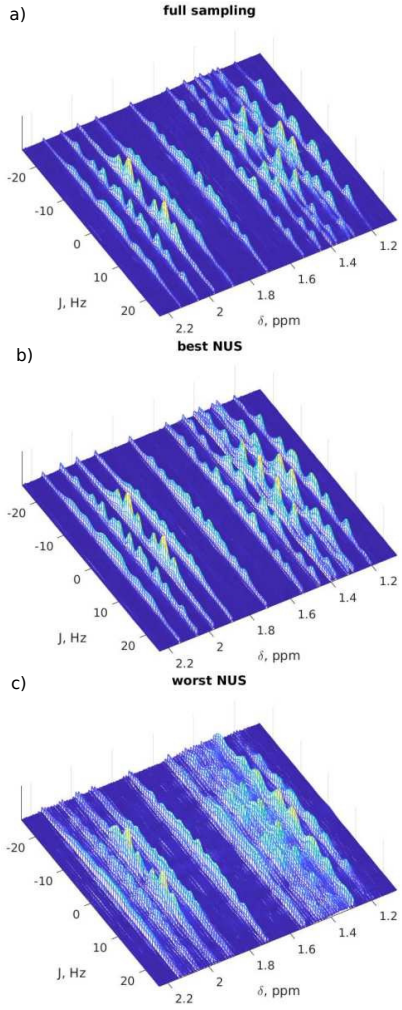

Fig. 3 Crowded region of a 2DJ-ZQS-PSYCHE spectrum of β-estradiol. a) Full sampling in F2, acquisition time - 7h 27min. b) and c) Chunks omitted in F2 according to the best (b) and worst (c) sampling (maximum criterion) and CS reconstruction. Acquisition time - 1h 16min.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Authors thank the National Science Centre of Poland for support with OPUS (2015/17/B/ST4/04221) grant.

Notes and references

1 K. Zangger, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2015, **86-87**, 1 – 20.
2 L. Castanar and T. Parella, *Magnetic Resonance in Chemistry*, 2015, **53**, 399–426.
3 I. E. Ndukwe, A. Shchukina, V. Zorin, C. Cobas, K. Kazimierczuk and C. P. Butts, *ChemPhysChem*, 2017, **18**, 2081–2087.
4 M. Foroozandeh, R. W. Adams, N. J. Meharry, D. Jeannerat, M. Nilsson and G. A. Morris, *Angewandte Chemie International Edition*, 2014, **53**, 6990–6992.
5 K. Kazimierczuk and V. Y. Orekhov, *Angewandte Chemie International Edition*, 2011, **50**, 5556–5559.
6 D. J. Holland, M. J. Bostock, L. F. Gladden and D. Nietlispach, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 6548–6551.
7 A. Shchukina, P. Kasprzak, R. Dass, M. Nowakowski and K. Kazimierczuk, *Journal of Biomolecular NMR*, 2017, **68**, 79–98.
8 H. Barkhuijsen, R. de Beer, W. M. Bovee, J. H. Creyghton and D. van Ormondt, *Magnetic Resonance in Medicine*, 1985, **2**, 86–89.
9 A. S. Stern and J. C. Hoch, *Magnetic Resonance in Chemistry*, 2015, **53**, 908–912.
10 S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhauser Basel, 2013.
11 I. E. Ndukwe, A. Shchukina, K. Kazimierczuk, C. Cobas and C. P. Butts, *ChemPhysChem*, 2016, **17**, 2799–2803.
12 I. E. Ndukwe, A. Shchukina, K. Kazimierczuk and C. P. Butts, *Chem. Commun.*, 2016, **52**, 12769–12772.
13 J. C. Barna, S. M. Tan and E. D. Lade, *Journal of Magnetic Resonance*, 1988, **78**, 327–332.
14 S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari and G. Wagner, *J. Biomol. NMR*, 2012, **52**, 315–327.
15 K. Kazimierczuk and V. Orekhov, *Magnetic Resonance in Chemistry*, 2015, **53**, 921–926.
16 M. Mobli and T. M. Miljenovic, *Journal of Magnetic Resonance*, 2019, **300**, 103 – 113.
17 S. G. Hyberts, K. Takeuchi and G. Wagner, *Journal of the American Chemical Society*, 2010, **132**, 2145–2147.
18 P. C. Aoto, R. B. Fenwick, G. J. Kroon and P. E. Wright, *Journal of Magnetic Resonance*, 2014, **246**, 31 – 35.
19 Y. Matsuki, M. T. Eddy and J. Herzfeld, *J. Am. Chem. Soc.*, 2009, **131**, 4648–4656.
20 J. J. Helmus and C. P. Jaroniec, *Journal of Biomolecular NMR*, 2013, **55**, 355–367.
21 V. Y. Orekhov, V. Jaravine, M. Mayzel and K. Kazimierczuk, 2004-2016.
22 I. Drori, *EURASIP Journal on Advances in Signal Processing*, 2007, **2007**, year.
23 S. G. Hyberts, S. A. Robson and G. Wagner, *Journal of Biomolecular NMR*, 2017, **68**, 139–154.
24 P. Kiraly, M. Foroozandeh, M. Nilsson and G. A. Morris, *Chemical Physics Letters*, 2017, **683**, 398 – 403.